

Appl. No. 09/848,982
Amendment and Response

NOV 30 2006

Docket No. 85804.014508

LISTING OF THE CLAIMS

1. (Currently Amended) A computer-implemented method of ~~text equivalence~~ ~~from a query string of characters~~ comprising:
modifying ~~the a~~ query string of characters using a predetermined set of heuristics;
performing a character-by-character comparison of the modified query string with
at least one known string of characters in a corpus in order to locate an exact match for the modified query string; and
responsive to not finding an exact match, performing the following steps ~~of in~~
order to locate an equivalent for the modified query string:
forming a plurality of sub-strings of characters from the query string, the
sub-strings having varying lengths such that at least two of the
formed sub-strings differ in length; and
using an information retrieval technique on the sub-strings formed from
the query string to identify a known string of characters equivalent
to the query string.
2. (Previously Presented) The method of claim 1, wherein the information retrieval
technique further comprises:
weighting the sub-strings;
scoring known strings of characters; and
retrieving information associated with a known string having the highest score.
3. (Previously Presented) The method of claim 2, further comprising, responsive to the
highest score being greater than a first threshold, automatically accepting the
known string having the highest score as an exact match.

Appl. No. 09/848,982
Amendment and Response

Docket No. 85804.014508

4. (Previously Presented) The method of claim 2, further comprising, responsive to the highest score being less than a second threshold and greater than a first threshold, presenting the known string having the highest score to a user for manual confirmation.
5. (Previously Presented) The method of claim 2, further comprising, responsive to the highest score being less than a second threshold and greater than a third threshold, presenting the known string having the highest score to a user to select the equivalent string.
6. (Previously Presented) The method of claim 1, wherein forming a plurality of sub-strings of characters comprises successively extending sub-strings based on frequency of occurrence in the modified query string.
7. (Previously Presented) The method of claim 1, wherein the query string is selected from the group consisting of a song title, a song artist, an album name, a book title, an author's name, a book publisher, a genetic sequence, and a computer program.
8. (Previously Presented) The method of claim 1, wherein the predetermined set of heuristics comprises removing whitespace from the query string.
9. (Previously Presented) The method of claim 1, wherein the predetermined set of heuristics comprises removing a portion of the query string.
10. (Previously Presented) The method of claim 1, wherein the predetermined set of heuristics comprises replacing a symbol in the query string with an alternate representation for the symbol.

Appl. No. 09/848,982
Amendment and Response

Docket No. 85804.014508

11. (Previously Presented) The method of claim 1 further comprising storing a database entry indicating that the query string is an equivalent of the identified known string.
12. (Currently amended) A computer-implemented system ~~for text equivalence from a query string of characters~~ comprising:
a heuristics module for modifying ~~the a~~ query string of characters using a predetermined set of heuristics;
a comparator module, coupled to the heuristics module, for performing a character-by-character comparison of the modified query string with at least one known string of characters in a corpus in order to find an exact match for the modified query string;
sub-string formation and information retrieval module responsive to not finding an exact match and for locating an equivalent for the modified query string;
said [[a]] sub-string formation module, coupled to the comparator module, ~~for, responsive to not finding an exact match,~~ forming a plurality of sub-strings of characters from the query string, the sub-strings having varying lengths such that at least two of the formed sub-strings differ in length; and
said [[an]] information retrieval module, coupled to the sub-string formation module, for performing an information retrieval technique on the sub-strings formed from the query string to identify a known string of characters equivalent to the query string.
13. (Previously Presented) The system of claim 12, wherein the information retrieval module further comprises:
a weight module for weighting the sub-strings;
a score module for scoring known strings of characters; and

Appl. No. 09/848,982
Amendment and Response

Docket No. 85804.014508

a retrieval module, coupled to the weight and score modules, for retrieving information associated with the known string having the highest score.

14. (Original) The system of claim 13, further comprising an accept module, coupled to the retrieval module, for accepting the information retrieved as an exact match for the highest score greater than a first threshold.
15. (Original) The system of claim 13, further comprising an accept module, coupled to the retrieval module, for presenting the information retrieved to a user for manual confirmation for the highest score less than a first threshold and greater than a second threshold.
16. (Original) The system of claim 13, further comprising an accept module, coupled to the retrieval module, for presenting the information retrieved to the user as a set of options for a user to select for the highest score less than a second threshold and greater than a third threshold.
17. (Previously Presented) The system of claim 12, wherein the sub-string formation module forms a plurality of substrings of characters by successively extending sub-strings based on frequency of occurrence in the modified query string.
18. (Previously Presented) The system of claim 12, wherein the query string is selected from the group consisting of a song title, a song artist, an album name, a book title, and author's name, a book publisher, a genetic sequence, and a computer program.
19. (Previously Presented) The system of claim 12, wherein the predetermined set of heuristics comprises removing whitespace from the query string.
20. (Previously Presented) The system of claim 12, wherein the heuristics module comprises a removal module for removing a portion of the query string.

Appl. No. 09/848,982
Amendment and Response

Docket No. 85804.014508

21. (Previously Presented) The system of claim 12, wherein the heuristics module comprises a replacement module for replacing a symbol in the query string with an alternate representation for the symbol.
22. (Previously Presented) The system of claim 12 further comprising a database update module for storing a database entry indicating that the query string is an equivalent of the identified known string.
23. (Currently amended) A computer-readable medium comprising computer-readable code ~~for performing text equivalencing from a query string of characters~~ comprising: computer-readable code adapted to modify ~~the~~ a query string of characters using a predetermined set of heuristics;
computer-readable code adapted to perform a character-by-character comparison of the modified query string with at least one known string of characters in a corpus in order to locate an exact match for the modified query string;
and
computer-readable code adapted to, responsive to not finding an exact match, locate an equivalent for the modified query string, comprising computer-readable code to:
form a plurality of sub-strings of characters from the query string, the sub-strings having varying lengths such that at least two of the formed sub-strings differ in length; and to
use an information retrieval technique on the sub-strings formed from the query string to identify a known string of characters equivalent to the query string.
24. (Previously Presented) The computer-readable medium of claim 23, wherein the information retrieval technique further comprises:
computer-readable code adapted to weight the sub-strings;

Appl. No. 09/848,982
Amendment and Response

Docket No. 85804.014508

computer-readable code adapted to score known strings of characters; and
computer-readable code adapted to retrieve information associated with a known
string having the highest score.

25. (Previously Presented) The computer-readable medium of claim 24, further comprising computer-readable code, responsive to the highest score being greater than a first threshold, adapted to automatically accept the known string having the highest score as an exact match.
26. (Previously Presented) The computer-readable medium of claim 24, further comprising computer-readable code, responsive to the highest score being less than a second threshold and greater than a first threshold, adapted to present the known string having the highest score to a user for manual confirmation.
27. (Previously Presented) The computer-readable medium of claim 24, further comprising computer-readable code, responsive to the highest score being less than a second threshold and greater than a third threshold, adapted to present the known string having the highest score to a user to select the equivalent string of characters.
28. (Previously Presented) The computer-readable medium of claim 23, wherein computer-readable code adapted to form a plurality of sub-strings of characters comprises computer-readable code adapted to successively extend sub-strings based on frequency of occurrence in the modified query string.
29. (Previously Presented) The computer-readable medium of claim 23, wherein the query string selected from a group consisting of a song title, a song artist, an album name, a book title, an author's name, a book publisher, a genetic sequence, and a computer program.

Appl. No. 09/848,982
Amendment and Response

Docket No. 85804.014508

30. (Previously Presented) The computer-readable medium of claim 23, wherein the predetermined set of heuristics comprises removing whitespace from the query string.
31. (Previously Presented) The computer-readable medium of claim 23, wherein the predetermined set of heuristics comprises removing a portion of the query string.
32. (Previously Presented) The method of claim 23, wherein the predetermined set of heuristics comprises replacing a symbol in the query string with an alternate representation for the symbol.
33. (Previously Presented) The computer-readable medium of claim 23 further comprising computer-readable code adapted to store a database entry indicating that the query is an equivalent of the identified known string.
34. (Currently Amended) A computer-implemented system ~~for performing text~~
~~equivalencing from a query string of characters~~ comprising:
a modifying means for modifying ~~the a~~ query string of characters using a predetermined set of heuristics;
a comparator means for performing a character-by-character comparison of the modified query string with at least one known string of characters in a corpus in order to locate an exact match for the modified query string;
formation and information retrieval means responsive to not finding an exact match and for locating an equivalent for the modified query string;
said [[a]] formation means ~~for, responsive to not finding an exact match,~~
forming a plurality of sub-strings of characters from the query string, the sub-strings having varying lengths such that at least two of the formed sub-strings differ in length; and

Appl. No. 09/848,982
Amondment and Response

Docket No. 85804.014508

said [[an]] information retrieval means operating on the sub-strings
formed from the query string for identifying a known string of
characters equivalent to the query string.

35. (Previously Presented) The system of claim 34, wherein the information retrieval means further comprises:
a weight means for weighting the sub-strings;
a score means for scoring known strings of characters; and
a retrieval means for retrieving information associated with the known string
having the highest score.
36. (New) The system of claim 35, wherein a weight for a given sub-string is based at least in part on a number of times the sub-string occurs in the query.
37. (New) The system of claim 34, wherein the length of a sub-string is determined based on one or more character sequences identified in the modified query string and a corresponding frequency of occurrence for each identified character sequences.
38. (New) The computer-implemented method of claim 1, wherein the length of a sub-string is determined based on one or more character sequences identified in the modified query string and a corresponding frequency of occurrence for each identified character sequences.
39. (New) The computer-implemented method of claim 2, wherein a weight for a given sub-string is based at least in part on a number of times the sub-string occurs in the query.
40. (New) The system of claim 12, wherein the length of a sub-string is determined based on one or more character sequences identified in the modified query string and a corresponding frequency of occurrence for each identified character sequences.

Appl. No. 09/848,982
Amendment and Response

Docket No. 85804.014508

41. (New) The system of claim 13, wherein a weight for a given sub-string is based at least in part on a number of times the sub-string occurs in the query.
42. (New) The computer-readable medium of claim 23, wherein the length of a sub-string is determined based on one or more character sequences identified in the modified query string and a corresponding frequency of occurrence for each identified character sequences.
43. (New) The computer-readable medium of claim 24, wherein a weight for a given sub-string is based at least in part on a number of times the sub-string occurs in the query.